Source:	TianHeng Shen(沈天珩, aka CheonHyeong Sim)
Title:	Disunification request for 8 characters
Status:	Individual Contribution on IRG #58
Action:	To be considered by IRG

Introduction

We found 8 pairs of characters that need to be disunified.

\>> U+681D（栝）

\>> U+6B25（步）

\>> U+6C97（汦）

\>> U+929B（銛）

\>> U+46B0（䚰）

\>> U+208B7（𠢷）

\>> U+216BA（妠）

\>> U+2335F（𣍟）

1. U+681D（栝）

According to the regional conventions in Taiwan, the right part of "活" is different from "舌", which can be distinguished by the first stroke. These two components are non−cognate. But U+681D has both the one with the right part "舌"(shé) and the another with the right part "舌"(guā), thus the two characters need to be disunified.

Figure 1　Evidence from 康熙字典同文書局原版 P522

栝 <small>説文·木部</small> 栝 <small>老子乙二二五上</small> 栝 <small>趙菿殘碑</small>

《説文》：“栝，炊竈木。从木，舌聲。”

（一）tiǎn 《廣韻》他玷切，上忝透。談部。

❶撥火棍。《説文·木部》：“栝，炊竈木。”桂馥義證：“炊竈木者，《六書故》：‘栝，進火木。’”段玉裁注：“炊竈木，今俗語云竈橇是也。《廣韻》云：‘橇，火杖。’栖（栝）、橇古今字也。”元尹廷高《車中作古乐府》：“停車少憩日又出，束栝營炊道傍屋。”

❷木杖。《集韻·忝韻》：“栝，木杖也。”唐李白《大獵賦》：“擺椿栝，開林叢。”《農政全書·農器·圖譜一》：“（耙）其桯兩端木栝，長可尺三，前稍微昂，穿兩木椆，以繫牛鞅鈎索。”

（二）kuò 《廣韻》古活切，入末見。月部。

❶樧。《集韻·末韻》：“栝，《説文》：‘樧也。’”

❷箭末扣弦处。《集韻·末韻》：“栝，矢栝築弦處。”《莊子·齊物論》：“其發若機栝，其司是非之謂也。”《淮南子·兵略》：“夫栝，淇衛箘簬，載以銀錫，雖有薄縞之幨，腐荷之繒，然猶不能獨射也。”高誘注：“栝，箭栝。”《樂府詩集·橫吹曲辭五·隔谷歌》：“弓無弦，箭無栝。”

❸木名，即桧。《廣雅·釋木》：“栝，栢也。”王念孫疏證：“栝，與檜同。《爾雅》云：‘檜，柏葉松身。’是栝即栢之別種，故以栝為栢也。”《玉篇·木部》：“栝，木名。栢葉松身。”《書·禹貢》：“杶榦栝柏。”孔傳：“柏葉松身曰栝。”唐杜甫《別張十三建封》：“雖當霜雪嚴，未覺栝柏枯。”

Figure 2　Evidence from 汉语大字典第二版 P1285

According to the evidences listed above, U+681D has two completely different pronunciations and meanings. The pronunciation tiǎn corresponds to the phonetic component “甜省”(i.e. 舌); while the pronunciation kuò corresponds to the phonetic component “舌”. So the two characters are phonetically, graphically and semantically different and need to be disunified.

681D
木 75.6

栝 栝 栝 栝 栝

G0-6869　HB2-D1EB　T2-2F6D　J13-755F　K1-596E

Figure 3　U+681D in CodeCharts of Unicode 14.0

In order not to affect the use of the regions where "舌" and "舌" are not distinguished, we suggest to retain kuò(i.e. 栝) on U+681D and encode tiǎn(i.e. 栝) separately.

Glyph	Mandarin	IDS	RS	FS	Variant	Source Reference
栝	tiǎn	⿰木舌	75.6	1	栎 U+3B87	GKX−0522.19

Table 1　Information of GKX−0522.19

2．U+6B25（歕）

The left part of U+6B25 is "曰"(yuē). It is also used for the another one with the left part "日", however, they are non−cognate.
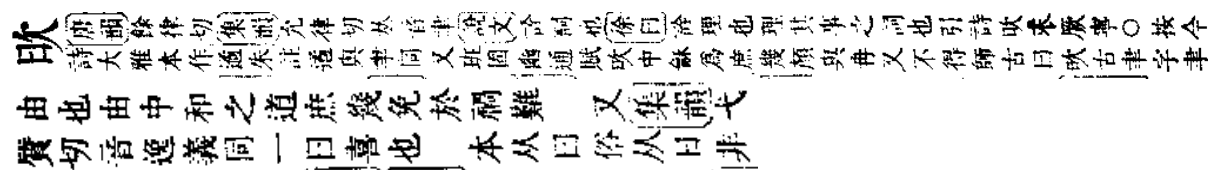

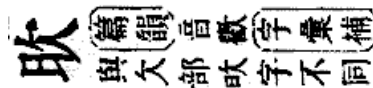
Figure 4　Evidence from 康熙字典同文書局原版 P566

Figure 5　Evidence from 康熙字典同文書局原版 P1600

Figure 6　Evidence from 文字情報基盤検索システム(JMJ−014753)

According to the evidences listed above, the phonetic component of U+6B25 is "曰"; while the semantic component of GKX−1600.50 is "日". So the two characters are phonetically, graphically and semantically different and need to be disunified.

Figure 7　U+6B25 in CodeCharts of Unicode 14.0

If GKX−1600.50 is separately encoded, we suggest China, TCA and ROK change the glyph of U+6B25 to distinguish one from another if necessary.

Glyph	Mandarin	IDS	RS	FS	Variant	Source Reference
昡	huān	⿱日欠	72.4	3	/	GKX−1600.15

Table 2　Information of GKX−1600.15

3. U+6C97（㳷）

GE−3F28, J14−6E3E and UTC−00763 is now unified with T3−2963 and MD−6C97. However, the one with the bottom part "氺" is non−cognate with the another one with the bottom part "水".
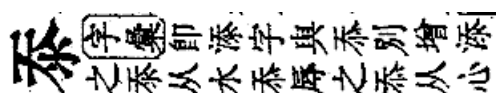

Figure 8　U+6C97 in CodeCharts of Unicode 14.0

Figure 9　Evidence from 康熙字典同文書局原版 P611

Figure 10　Evidence from 康熙字典同文書局原版 P1592

㳷 pāng 《字彙補·大部》："㳷，坡雙切。見《海篇》。"

㳷 (一) tiān

同"添"。增加。《字彙·水部》："㳷，即添字。與忝字不同，增㳷之㳷从水；忝辱之忝从心。"

Figure 11   Evidence from 汉语大字典第二版 P1667

Figure 12   Evidence from 文字情報基盤検索システム(JMJ−015148)

Figure 13   Evidence from 文字情報基盤検索システム(JMJ−015147)

According to the evidences listed above, the two characters are phonetically, graphically and semantically different and need to be disunified. By the way, this character can be horizontally extended by J−source as JMJ−015147 if necessary. In addition, if this character is separately encoded, the M−source should probably change MD−6C97 to MD−XXXXX(XXXXX for the new codepoint).

Glyph	Mandarin	IDS	RS	FS	Variant	Source Reference
泰	pāng	⿱天水	37.5	1	滂 U+6EC2	GKX−1592.28 T3−2963 MD−6C97 (JMJ−015147)

Table 3   Information of GKX−1592.28

4. U+929B（銛)

According to the regional conventions in Taiwan, the right part of "活" is different from "舌", which can be distinguished by the first stroke. These two components are non−cognate. But U+929B has both the one with the right part "舌"(shé) and the another with the right part "舌"(guā), thus the two characters need to be disunified.
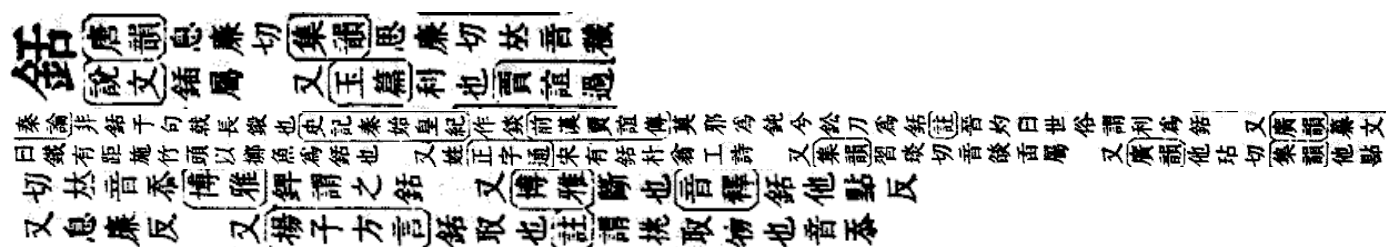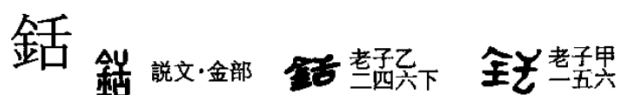

Figure 14　Evidence from 康熙字典同文書局原版 P1305

Figure 15　Evidence from 汉语大字典第二版 P4520

According to the evidences listed above, U+929B has two completely different pronunciations and meanings. The pronunciation xiān/tiǎn correspond to the phonetic component "甜省"(i.e. 舌); while the pronunciation guā corresponds to the phonetic component "舌". So the two characters are phonetically, graphically and semantically different and need to be disunified.
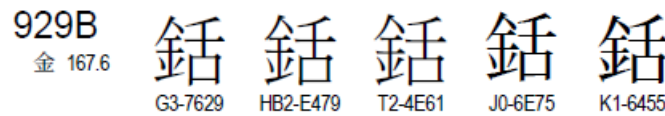

Figure 16　U+929B in CodeCharts of Unicode 14.0

In order not to affect the use of the regions where "舌" and "舌" are not distinguised, we suggest to retain guā(i.e. 銛) on U+929B and encode xiān/tiǎn(i.e. 銛) separately.

Glyph	Mandarin	IDS	RS	FS	Variant	Source Reference
銛	xiān tiǎn	⿰金舌	167.6	1	錟 U+931F	GKX−1305.11

Table 4　Information of GKX−1305.11

5. U+46B0（訮）

G5−7152 is now unified with T3−2F2D. However, the one with the right part "匕" is non−cognate with the another one with the right part "匕". These two characters have the same case as "叱" and "叱", which was mentioned in IRGN2447 by Kushim Jiang(姜兆勤) and was already disunified in September, 2021 along with the official release of Unicode 14.0.

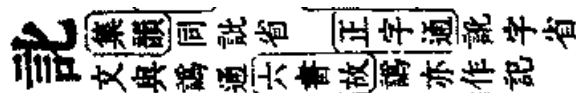Figure 17　U+46B0 in CodeCharts of Unicode 14.0

Figure 18　Evidence from 康熙字典同文書局原版 P1146

訛 (一)pǐ 《集韻》普弭切，上紙滂。
同"訛"。言具。《集韻·紙韻》："訛，言具也。或省。"
(二)é

同"訛"。《正字通·言部》："訛，訛字省文。"

Figure 19 Evidence from 汉语大字典第二版 PP4195−4196

Figure 20 Evidence from TCA−CNS 11643(T3−2F2D)

Figure 21−22　Evidence from 教育部異體字字典(B00951, A03801)

Figure 23　Evidence from 文字情報基盤検索システム(JMJ−004545)

Figure 24　Evidence from 文字情報基盤検索システム(JMJ−004544)

According to the evidences listed above, the two characters are phonetically, graphically and semantically different and need to be disunified. By the way, this character can be horizontally extended by J−source as JMJ−004544 if necessary.

Glyph	Mandarin	IDS	RS	FS	Variant	Source Reference
訛	é	⬚言匕	149.2	1	訛 U+8A1B	GKX−1146.14 T3−2F2D *(JMJ−004544)*

Table 5　Information of T3−2F2D

6. U+208B7（�）

It is widely known that "来" is the simplification form of "來" and shouldn't be unified. But GKX−0149.39 is unified with T5−5A7C, that seems to be strange.

Figure 25　U+208B7 in CodeCharts of Unicode 14.0

In CNS 11643, only "�" exists; in 教育部異體字字典, both "勲" and "勲" exist, but all the documentary evidences listed on the page show the glyph "勲" while no one shows "勲". Thus we request TCA to give the evidence of "勲", and disunify these two characters; or TCA should probably change the glyph in order to match the evidences.

Glyph	Mandarin	IDS	RS	FS	Variant	Source Reference
勲	kǔn	⬚勑困	19.15	1	闃 U+95AB	T5−5A7C

Table 6　Information of T5−5A7C(if disunify)

7. U+216BA（妠）

According to the regional conventions in Hong Kong and Taiwan, the right part of "執" is different from "丸", which can be distinguished by the relative position of the first stroke and the last stroke. These two components are non−cognate. This is also mentioned in the Appendix A of IRGN2225. But U+216BA has both the one with the right part "丸"(wán) and the another with the right part "丸"(jǐ), thus the two characters need to be disunified.

Figure 26−27　U+216BA in CodeCharts of Unicode 14.0/5.2

Appendix A: Rationale for Inclusion of Z3-03

Z3-03 丸 is included as a separate component from U+4E38 丸. First, they are considered different components in the Taiwan and Hong Kong locale. For characters whose etymology is U+4E2E 丮 (to hold), the glyph shape Z3-03 with the dot protruding is used. For characters whose etymology is U+4E38 丸 (small round object), the T/H glyph shape of U+4E38 with the dot not protruding is used.

Figure 28   U+216BA in the Appendix A of IRGN2225

In both 康熙字典 and 汉语大字典, this character is interpreted as the variation of "㚲", which has the pronunciation fàn and the component "丸"(wán). In HKSCS, it is read jyun4 in Cantonese, which is as same as "丸". But in CNS 11643, it gives the pronunciations fàn and zhí, while the latter is the variation of "執" and the right component is "丮"(jǐ). The T−source glyph in Unicode 5.2 is "㚲"(fàn), but since Unicode 6.0, it has changed to "�...丮"(zhí).

Figure 29   Evidence from TCA−CNS 11643(T6−2443)

We can see that 明體 has the component "丸"(wán) while 楷體 and 宋體 have the component "丮"(jǐ).

Figure 30 Evidence from 教育部異體字字典(A00784)

In 教育部異體字字典, the evidence of "奻"(zhí) refers to 古文四聲韻. Below is the evidence from 古文四聲韻.

Figure 31 Evidence from 欽定四庫全書 古文四聲韻 卷五 P51

The evidence shows that this character has the same vowel as "緝" in Ancient Chinese, in other words, its Mandarin pronunciation should be zhí. Thus, we suggest TCA consider restoring the glyph on U+216BA to the one in Unicode 5.2, and disunifying "妷" from "妷".

Glyph	Mandarin	IDS	RS	FS	Variant	Source Reference
妷	zhí	⿰女丸	38.3	3	執 U+57F7	T13−XXXX

Table 7 Information of 妷(zhí)

8. U+2335F（冎）

U+2F980 冎 is unified with U+2335F 冎 according to [WG2 N2159R](). However, according to the evidences from G−source and T−source, the two characters need to be disunified. U+2335F has the bottom component 月 and the pronunciation wěi, while U+2F980 has the bottom component 肉 and the pronunciation kěn. Disunifying the characters with the opposition of 月 and 肉 also appeared in [IRGN2239](), and following 4 characters was successfully disunified in March, 2020 along with the official release of Unicode 13.0.

Char.	Unicode	Radical	Char.	Unicode	Radical
朒	U+4DBA	月	朒	U+80AD	肉
肬	U+4DBB	月	肬	U+43D9	肉
膆	U+4DBC	月	膆	U+440B	肉

| 腜 | U+4DBD | 月 | 腜 | U+6721 | 肉 |

This character has the same case as the four characters above.

Figure 32　Evidence from 康熙字典同文書局原版 P504

Figure 33　Evidence from 康熙字典同文書局原版 P973

Figure 34　Evidence from 康熙字典同文書局原版 P976

The glyph in P504 has the radical 月, the right side of the two strokes "一" in the component "月" do not contact the stroke "丁"; while the glyphs in P973 and P976 have the radical 肉, the right side of the two strokes "一" in the component "月" do contact the stroke "丁". This method is commonly used in 康熙字典 in order to distinguish Radical 月 and Radical 肉.

Figure 35　Evidence from 汉语大字典第二版 P2194

Although it says "説文・肉部" in the total start, "改併四聲篇海・月部" is written when interpreting the character with the pronunciation wěi.

Figure 36　Evidence from TCA−CNS 11643(T5−2444)
Figure 37　Evidence from TCA−CNS 11643(T6−287B)

Figure 38−39　Evidence from 教育部異體字字典(C04968, A03307)

Figure 40　Evidence from 文字情報基盤検索システム(JMJ−037876)

Figure 41　Evidence from 文字情報基盤検索システム(JMJ−037877)

We can see both the Japanese pronunciation and the radical of the two characters are completely different in 文字情報基盤検索システム; in addition, these two characters are also separated in 大漢和辞典, with the serial numbers 14334 and 29258 respectively.

Figure 42−43　U+216BA/U+2F980 in CodeCharts of Unicode 14.0

According to the evidences listed above, the two characters are phonetically, graphically and semantically different and need to be disunified. By the way, this character can be horizontally extended by G−source as GKX−0973.20 and J−source as JMJ−037877 if necessary. In addition, the T−source of U+2F980 needs to change to TU−2F980 and the radical needs to change to 肉.

Glyph	Mandarin	IDS	RS	FS	Variant	Source Reference
冐	kěn	⿱二月	130.3	4	肯 U+80AF	GKX−0973.20 T6−2F2D *(JMJ−037877)*

Table 9　Information of GKX−0973.20

(End of document)